Control-theoretic Approach to Communication with Feedback: Fundamental Limits and Code Design

Ehsan Ardestanizadeh, *Student Member, IEEE* and Massimo Franceschetti, *Member, IEEE*

arXiv:1006.5265v1 [cs.IT] 28 Jun 2010

Abstract—**Feedback communication is studied from a control-theoretic perspective, mapping the communication problem to a control problem in which the control signal is received through the same noisy channel as in the communication problem, and the (nonlinear and time-varying) dynamics of the system determine a subclass of encoders available at the transmitter. The *MMSE capacity* is defined to be the supremum exponential decay rate of the mean square decoding error. This is upper bounded by the information-theoretic feedback capacity, which is the supremum of the achievable rates. A sufficient condition is provided under which the upper bound holds with equality. For the special class of stationary Gaussian channels, a simple application of Bode's integral formula shows that the feedback capacity, recently characterized by Kim, is equal to the maximum *instability* that can be tolerated by the controller under a given power constraint. Finally, the control mapping is generalized to the N-sender AWGN multiple access channel. It is shown that Kramer's code for this channel, which is known to be sum rate optimal in the class of generalized linear feedback codes, can be obtained by solving a linear quadratic Gaussian control problem.**

I. INTRODUCTION

Feedback loops are central to many engineering systems. Their study naturally falls at the intersection between communication and control theories. However, the information-theoretic approach and the control-theoretic one have often evolved in isolation, separated by almost philosophical differences. In this paper we attempt one step at bridging the gap, showing how tools from both disciplines can be applied to study fundamental limits of feedback systems and to design efficient codes for communication in the presence of feedback.

Consider the feedback communication problem over an arbitrary point-to-point channel depicted in Fig 1a. The encoder, which has access to the channel outputs causally and noiselessly, wishes to communicate a continuous message $M \in (0, 1)$ to the decoder through n channel uses. At the end of the transmissions, the decoder forms an estimate $\hat{M}$ based on the received channel outputs, and the *mean square error* (MSE) of the estimate $\hat{M}$ represents the performance metric of the communication.

We map this communication problem to the general (*nonlinear and time varying*) controlled dynamical system depicted in Fig. 1b, in which the initial state of the system corresponds to the message M, and the control actions–received through the same noisy channel as in the communication problem– correspond to the transmitted signals by the encoder. In this representation, the set of controllers for a given system

E. Ardestanizadeh, M. Franceschetti are with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA, 92093-0407, email: ehsan@ucsd.edu, massimo@ece.ucsd.edu.

corresponds to a subclass of encoders for the communication problem. In fact, the system can be viewed as a filter which determines the *information pattern* [1], on which the transmitted signals (actions) by the encoder (controller) can depend upon. A similar mapping for the special case of *linear time-invariant* (LTI) systems and controllers was first presented in [2].

We study three different channel models. First, we consider a general point-to-point channel. The *MSE exponent* is defined as the exponential decay rate of the MSE in block length n, and the (feedback) *minimum mean square error (MMSE) capacity* is defined as the supremum of all achievable MSE exponents with feedback. We show that the MMSE capacity is upper bounded by the information-theoretic (feedback) capacity, the supremum of all achievable rates with feedback. We also present a sufficient condition, under which the (information-theoretic) capacity coincides with the MMSE capacity. These results provide a step towards the understanding of the connection between estimation and information theory.

Second, we focus on the stationary Gaussian channel with feedback, the capacity of which was recently characterized by Kim [3]. Applying the discrete extension of Bode's result [4] (cf. [5], [6]), we observe that the capacity of the Gaussian channel under power constraint P is equal to the maximum instability which can be tolerated by a linear controller with power at most P, acting over the same stationary Gaussian channel. This follows almost immediately from the previous results [2], [3] and provides a step towards the understanding of the connection between stabilizability over some noisy channel and the capacity of that channel.

Finally, we consider the N-sender additive white Gaussian noise (AWGN) multiple access channel (MAC) with feedback depicted in Fig. 2a. We show that the linear code proposed by Kramer [7], which is known to be optimal among the class of generalized linear feedback codes [8], can be obtained as the optimal solution of a linear quadratic Gaussian (LQG) problem given by a linear time-invariant (LTI) system controlled over a point-to-point AWGN channel where the asymptotic cost is the average power of the controller. These results provide a step towards the understanding of how control tools can be used to design codes for communication.

We now wish to place our results in the context of the related literature. The results on the N-sender AWGN-MAC generalize previous ones of Elia [2], who recovers Ozarow's code [9]–a special case of Kramer's code–using control theory. Our approach is different from Elia's in both the model and the analysis. Our reduction is to a control problem over a single point-to-point channel for any $N \geq 2$ number of senders, and our analysis is based on the theory of LQG control. In

contrast, in [2] the communication is limited to 2-sender MAC, which is mapped to a control problem also over a 2-sender MAC, and the analysis is based on the technique of Youla parameterization.

The connection between the MMSE and capacity has been investigated extensively and from different perspectives in the literature. For example, in a classic paper Duncan [10] expresses the mutual information between a continuous random process and its noisy version corrupted by white noise, in terms of the causal MMSE. More recently, Forney [11] explains the role of the MMSE in the context of capacity achieving lattice codes over AWGN channels. Guo et al. [12] and Zakai [13] showed that for a discrete random vector observed through an AWGN channel, the derivative of the mutual information between input and output sequences with respect to the signal-to-noise ratio (SNR), is half the (noncausal) MMSE. We point out that these authors study the average MMSE of a vector observed over a noisy channel without feedback as a function of the SNR. In contrast, we consider the estimation of a single random variable (message), given the observation of a whole block of length n, and we look at the exponential decay rate of the MMSE with n, at fixed SNR. Of more relevance to us is the recent work of Liu and Elia [14], who study linear codes over Gaussian channels obtained using a Kalman filter (KF) approach. For this class of codes, they show that the decay rate of the MMSE equals the mutual information between the message and the output sequence. In contrast, our results for the MMSE capacity are derived based on an information-theoretic approach and hold for all codes over general channels.

Additional works in the literature revealed connections between control theory and information theory. Without attempting of being exhaustive, we distinguish between those who use information theory to study control systems and those who use control theory to study communication systems. Within the first group, Mitter and Newton [15], [16] studied estimation and filtering in terms of information and entropy flows. In the last decade, Bode-like fundamental limitations in controlled systems have been analyzed with success from an information-theoretic perspective [17], [18], [19], [20], [21]. In this context, we point out that our Lemma 3, when it is specialized to the case of additive channels, provides an alternative proof of Theorem 4.2 in [20].

Within the second group, Elia [2] was the first to map linear codes for additive white Gaussian noise channels to an LTI system controlled over an AWGN channel. Subsequently, Wu et al. [22] studied the Gaussian interference channel in terms of estimation and control. Tatikonda and Mitter [23] used a Markov decision problem (MDP) formulation to study the capacity of Markov channels with feedback, and recently Coleman [24] considered the design of the feedback encoder from a stochastic control perspective. Finally, we also refer the reader to the work in [25], which gives an historical perspective and contains selected additional references.

The rest of the paper is organized as follows. Section II presents the definitions and the mapping between the feedback communication and the control problem. Section III provides the upper bound on the MMSE capacity for a general point-to-point channel. The point-to-point stationary Gaussian channels and the AWGN multiple access channel are considered in Section IV and Section V, respectively. Finally, Section VI concludes the paper.

Notation: A random variable is denoted by an upper case letter (e.g. X, Y, Z) and its realization is denoted by a lower case letter (e.g. x, y, z). Similarly, a random column vector and its realization are denoted by bold face symbols (e.g. $\mathbf{X}$ and $\mathbf{x}$, respectively). Uppercase letters (e.g. A, B, C) also denote matrices, which can be differentiated from a random variable based on the context. The (i, j)-th element of A is denoted by A_{ij}, and notation A^T and A' denote the transpose and complex transpose of matrix A, respectively. We use the following short notation for covariance matrices $K_{\mathbf{XY}} := \mathsf{E}(\mathbf{XY}') - \mathsf{E}(\mathbf{X})\,\mathsf{E}(\mathbf{Y}')$ and $K_{\mathbf{X}} := K_{\mathbf{XX}}$.

II. DEFINITIONS AND CONTROL APPROACH

Consider the communication problem depicted in Fig. 1a, where the sender wishes to convey a message $M \in \mathcal{M} := (0, 1)$ to the receiver through n uses of a stochastic channel,

$$Y_i = h_i(X_i, Z_i), \quad i = 1, \ldots, n.$$

where $X_i \in \mathcal{X}$ and $Y_i \in \mathcal{Y}$ denote the input and output of the channel, respectively, and $Z_i \in \mathcal{Z}$ denote the noise at time i. The set of mappings

$$h_i : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}, \quad i = 1, \ldots, n$$

and the distribution of the noise sequence $\{Z_i\}_{i=1}^n$ determine the channel. The noise process $\{Z_i\}$ is assumed to be independent of the message M. We assume that the output symbols are causally fed back to the sender and the transmitted symbol X_i at time i can depend on both the previous channel output sequence $Y^{i-1} := \{Y_1, Y_2, ..., Y_{i-1}\}$ and the message M.

Definition 1: We define a (feedback) n-code as

1) an encoder: a set of (stochastic)[1] encoding maps $f_i : \mathcal{M} \times \mathcal{Y}^{i-1} \to \mathcal{Y}$, also known to the receiver, such that for each $i = 1, \ldots, n$

$$X_i = f_i(M, Y^{i-1}) \tag{1}$$

   and

2) a decoder: a decoding map $\phi : \mathcal{Y}^n \to \mathcal{M}$ which determines the estimate of the message $\hat{M}$ based on the received sequence Y^n, i.e.,

$$\hat{M} = \phi(Y^n). \tag{2}$$

We assume that the message $M \in (0, 1)$ is a random variable uniformly distributed over the unit interval and does not depend on n. As the performance measure of the communication, we consider the MSE,

$$D^{(n)} := \mathsf{E}\left((M - \phi(Y^n))^2\right). \tag{3}$$

where the expectation is with respect to randomness of both the message and the channel. Note that the decoder does not affect the joint distribution of (M, X^n, Y^n, Z^n) and simply

---

[1]For stochastic encoders we can write X_i as a function of (M, Y^{i-1}, V_i), where $\{V_i\}$ is a random process independent of M and $\{Z_i\}$.
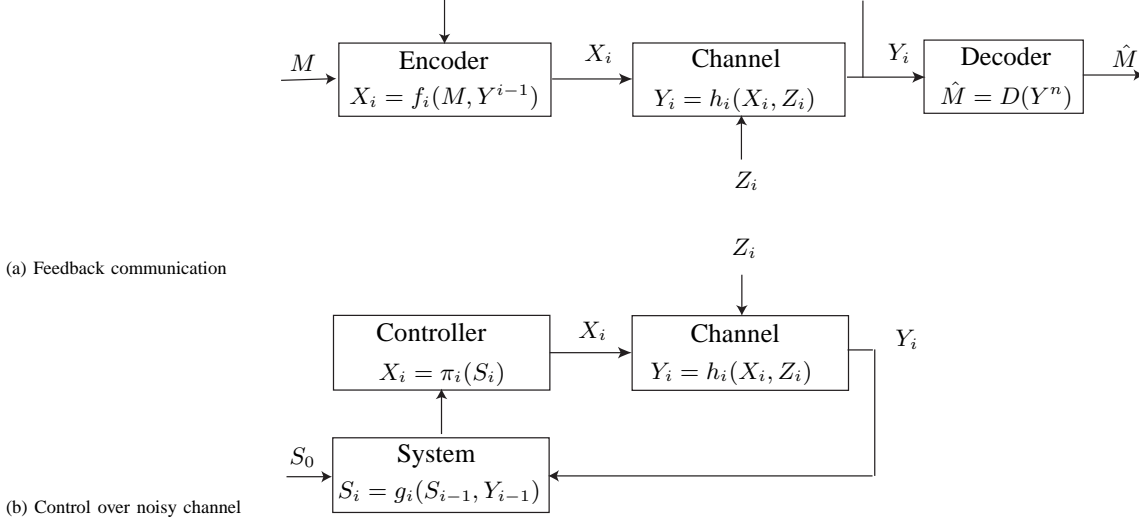
Fig. 1. An arbitrary encoder for the feedback communication over a stochastic channel depicted in (a) can be represented as a dynamical system controlled over the same channel depicted in (b).

estimates the message at the end of the block. Hence, without loss of generality, we pick the optimal decoder, namely, the MMSE estimator of the message given Y^n, and we call an encoder optimal if it minimizes $D^{(n)}$. Let

$$E^{(n)} := -\frac{1}{2n} \log(D^{(n)})$$

be the exponential decay rate of the MSE with respect to n.

Definition 2: The *MSE exponent* E is called achievable (with feedback) if there exists a sequence of n-codes such that

$$E \le \liminf_{n \to \infty} E^{(n)}. \tag{4}$$

The *MMSE capacity* C_{MMSE} is the supremum of all achievable MSE exponents.

The communication problem described above can be viewed as a control problem where the encoder wishes to control the knowledge of the receiver about the message. In his notable paper [26], Witsenhausen wrote: "When communication problems are considered as control problems (which they are), the information pattern is never classical since at least two stations, not having access to the same data, are always involved". However, this is not the case for the feedback communication problem in Fig. 1a since the encoder (controller) has access also to the information available at the decoder via feedback. In fact, below we show that this feedback communication problem can be represented as a control problem in which the controller has the complete state information.

Consider the control problem in Fig. 1b where the state at time i is

$$S_i = g_i(S_{i-1}, Y_{i-1}), \quad i = 1, \ldots, n \tag{5}$$

with initial state

$$S_0 = M$$

and $Y_0 = \emptyset$. We refer to the mappings $\{g_i\}_{i=1}^n$,

$$g_i : \mathcal{S}_{i-1} \times \mathcal{Y} \to \mathcal{S}_i, \qquad i = 1, \ldots, n$$

as the *system*. The controller, which observes the current state S_i, picks an action (symbol) $X_i \in \mathcal{X}$,

$$X_i = \pi_i(S_i), \quad i = 1, \ldots, n \tag{6}$$

according to a set of (stochastic) mappings

$$\pi_i : \mathcal{S}_i \to \mathcal{X}, \quad i = 1, \ldots, n.$$

We refer to the set $\{\pi_i\}_{i=1}^n$ as the *controller*.

The communication problem in Fig. 1a can be represented as the control problem Fig. 1b as follows. Let the system $\{g_i\}_{i=1}^n$ be such that the state S_i at time i is the collection of the initial state $S_0 = M$ and the past observations Y^{i-1}, namely,

$$S_i = (M, Y^{i-1}) \quad i = 1, \ldots, n. \tag{7}$$

Also, let the controller $\{\pi_i\}_{i=1}^n$ be picked according to the encoder $\{f_i\}_{i=1}^n$ in the communication problem such that

$$\pi_i(S_i) = f_i(M, Y^{i-1}), \qquad i = 1, \ldots, n.$$

Then the joint distribution of all the random variables (M, X^n, Y^n, Z^n) in the control problem is the same as that in the communication problem.

To complete the representation, let $\hat{S}_0(Y^n)$ be the MMSE estimate of the initial state S_0 based on Y^n, and the final cost be

$$c_n(S_n) := (S_0 - \hat{S}_0(Y^n))^2.$$

We call a controller optimal if it minimizes the final expected cost

$$\mathsf{E}(c_n(S_n)) = \mathsf{E}\left((M - \hat{M}(Y^n))^2\right)$$
$$= D^{(n)}.$$

Thus, the optimal controller represents the optimal encoder for the communication problem.

The system in (7) is the most general system which can represent all the encoders for the communication system. However, if the system $\{g_i\}_{i=1}^n$ is more restricted such that the state S_i is a filtered version of (S_0, Y^{i-1}), then the controller $\{\pi_i\}_{i=1}^n$ represents only a subclass of encoders $\{f_i\}_{i=1}^n$ where the transmitted symbol X_i depends on (M, Y^{i-1}) only through S_i (see (6)). In that case, we can view the system as a *filter* which determines the *information pattern* available at the controller (encoder). Below, we show a subclass of encoders for which the state S_i does not include all the past output Y^{i-1} as in (7), yet it contains all the optimal encoders. Let $F_{M|Y^{i-1}}(\cdot|Y^{i-1})$ be the conditional distribution of the message M given the channel outputs Y^{i-1}.

Lemma 1: An optimal encoder which minimizes the MSE $D^{(n)}$, can be found in the subclass of encoders which is determined by the system with state of the form $S_i = (M, F_{M|Y^{i-1}}(\cdot|Y^{i-1}))$.

   Proof: See Appendix A.

This lemma, which is based on a well-known result in stochastic control, provides a sufficient information pattern such that the optimal feedback encoders for the communication over a general channel can be built upon. For example, in the special case of memoryless channels, Shayevitz and Feder [27] proposed an explicit encoder which uses the information pattern described in Lemma 1, and showed that it is optimal in terms of the achievable rates.

III. MMSE CAPACITY

In this section we present the relationship between the MMSE capacity and the information-theoretic capacity [28].

First, we review the definition of the information-theoretic achievable rates, which is an asymptotic measure based on a sequence of message sets whose size depend on the block length n. Consider the n-code in Definition 1 and let the message $M_n \sim \text{Unif}(\mathcal{M}_n)$ be uniformly distributed over the set $\mathcal{M}_n := \{1, 2, \ldots, 2^{nR}\}$ such that

$$X_i = f_i(M_n, Y^{i-1}), \ i = 1, \ldots, n. \tag{8}$$

Let probability of error be $P_e^{(n)} = \mathsf{P}(M_n \neq \hat{M}_n)$. A rate R is called achievable (with feedback) if there exists a sequence of n-codes with message sets $\mathcal{M}_n$ for which $P_e^{(n)} \to 0$ as $n \to \infty$. The (feedback) capacity C is the supremum of all achievable rates.

Recall that the MSE exponent E is defined based on a message $M \in (0, 1)$ which does not depend on the block length n. The following lemma provides the connection between the achievable rates and MSE exponents.

Lemma 2: If the MSE exponent E is achievable, then any rate $R < E$ is achievable. Conversely, if the rate R is achievable and the probability of error satisfies $\lim_{n\to\infty} \frac{P_e^{(n)}}{2^{-2nR}} = 0$, then the MSE exponent $E = R$ is achievable.

   Proof: See Appendix B.

According to Lemma 2, showing that an MSE exponent $E > R$ is achievable for a uniform message $M \in (0, 1)$ is sufficient to show that rate R is achievable.

Below, we provide a general upper bound on the MMSE capacity. Before proceeding, we need some information-theoretic definitions [28]. Let $h(X)$ denote the differential entropy of a random variable X and $I(X; Y)$ be the mutual information between random variables X and Y. Moreover, let the *minimum entropy rate* of a general random process $\{X_i\}_{i=1}^\infty$ be defined as

$$\underline{h}(X) := \liminf_{n \to \infty} \frac{h(X^n)}{n}.$$

For a stationary random process, the limit exists [28] and the *entropy rate* is defined as

$$\bar{h}(X) := \lim_{n \to \infty} \frac{h(X^n)}{n}.$$

We call a channel invertible if Z_i can be recovered from (X_i, Y_i), namely, for every time $i > 0$ there exists a function h_i^{-1} such that

$$Z_i = h_i^{-1}(X_i, Y_i).$$

Lemma 3: For a system controlled over an invertible channel, we have

$$I(S_0; Y^n) \leq h(Y^n) - h(Z^n)$$

where equality holds if and only if X_i is a deterministic function of (S_0, Y^{i-1}) for $i = 1, \ldots, n$.

   Proof: See Appendix C.

Remark 1: Theorem 4.2 in [20] can be recovered from Lemma 3 by considering the special case of additive channels $Y_i = X_i + Z_i$ which are invertible.

Theorem 1: For an invertible channel, the MMSE capacity C_{MMSE} is upper bounded as

$$C_{MMSE} \leq C \leq \sup \ \underline{h}(Y) - \underline{h}(Z).$$

where the supremum is over all sequences of n-codes.

Remark 2: Theorem 1 holds without any assumption on the noise sequence or the controlled system.

   Proof:
   The first inequality

$$C_{MMSE} \leq C \tag{9}$$

can be shown using Lemma 2. Suppose (9) does not hold and $C < C_{MMSE}$. Then, we can always find an achievable MSE exponent E^* such that $C < E^*$, and subsequently we can pick a rate R^* such that $C < R^* < E^*$. Hence, by Lemma 2 the rate R^* is achievable, and this contradicts the definition of C. Therefore, the inequality (9) must hold.

For the second inequality, from Fano's inequality we have

$$R \leq \frac{1}{n} I(S_0, Y^n) + \epsilon_n.$$

where $\epsilon_n \to 0$ as $n \to \infty$. Hence, for a given sequence of n-codes

$$R \leq \liminf_{n \to \infty} \frac{1}{n} I(S_0, Y^n). \tag{10}$$

From Lemma 3 we have $h(Z^n) + I(S_0; Y^n) \leq h(Y^n)$. Taking the limit and rearranging terms we get

$$\liminf_{n \to \infty} \frac{1}{n} I(S_0, Y^n) \leq \underline{h}(Y) - \underline{h}(Z).$$

Combining with (10) we have $R \leq \underline{h}(Y) - \underline{h}(Z)$, and taking the supremum over all sequences of n-codes completes the proof. ■

Theorem 1 shows that in general, (feedback) MMSE capacity is upper bounded by the (feedback) capacity. Below, we provide a sufficient condition on the channel under which this bound is tight.

Corollary 1: For a given channel, we have

$$C_{MMSE} = C$$

if any rate $R < C$ can be achieved such that

$$\lim_{n \to \infty} \frac{P_e^{(n)}}{2^{-2nC}} \to 0. \qquad (11)$$

Proof: By assumption for any rate $R < C$, the condition on the decay rate of probability of error $P_e^{(n)}$ in Lemma 2 is satisfied and hence any MSE exponent $E = R < C$ is achievable. Therefore, we conclude that $C_{MMSE} \geq C$. On the other hand, by Theorem 1 we have $C_{MMSE} \leq C$, and hence $C_{MMSE} = C$. ■

In the next section we show that this equality holds for some special cases of Gaussian channels including the AWGN channel.

IV. POINT-TO-POINT GAUSSIAN CHANNELS

In this section, we turn our attention to additive Gaussian channels,

$$Y_i = X_i + Z_i$$

where the noise sequence $\{Z_i\}$ is a (colored) stationary Gaussian process with power spectral density $S_Z(e^{j\omega})$. The transmitted symbols are assumed to satisfy the (block) power constraint P, i.e.,

$$\frac{1}{n} \sum_{i=1}^{n} \mathsf{E}(X_i^2) \leq P.$$

We define the subclass of *LTI encoders* to be the encoders which can be represented by an LTI system and a linear controller in Fig. 1b. In this case, the system and the controller combined can be represented as as a *causal* filter

$$F(z) = \frac{X(z)}{Y(z)}$$

where $X(z)$ and $Y(z)$ are the Z-transform of the input and the output sequence, respectively. We alternatively refer to $F(z)$ as the open loop transfer function. Let the instability I be

$$I = \sum_{j=1}^{m} \log(|\beta_j|).$$

where β_j are the m unstable poles of the open loop transfer function $F(z)$.

Theorem 2: Under power constraint P, the capacity of the stationary Gaussian channel is

$$C = \sup \ I$$

where the supremum is over all causal filters $F(z)$ which satisfy the power constraint

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{F(e^{j\omega})}{1 - F(e^{j\omega})} \right|^2 S_Z(e^{j\omega}) d\omega \leq P. \qquad (12)$$

Remark 3: Theorem 2 links the two works and is a simple consequence of [2], [3]. The Gaussian channel capacity can be represented as the limit of a sequence of optimization problems [29]. Kim [3] proved that the limit of this sequence can be replaced by a single optimization problem, maximizing the entropy rate of the output process over all stationary Gaussian input processes, which are a causal and linear filtered version of the noise process. On the other hand, Elia [2] expressed the asymptotic average directed information between the input and output process for a stationary linear scheme over a Gaussian channel in terms of the Bode integral. We connect the capacity, the supremum achievable rate of communication, to the supremum tolerable instability over Gaussian channels.

Proof: The capacity of the stationary Gaussian channel under power constraint P is given by [3, Theorem 4.6]

$$C = \sup_{B} \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \left(|1 + B(e^{j\omega})|^2 \right) d\omega \qquad (13)$$

where the maximization is over all causal filters $B(e^{j\omega})$ which satisfy

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |B(e^{j\omega})|^2 S_Z(e^{j\omega}) d\omega \leq P. \qquad (14)$$

This characterization can also be viewed as follows [3],

$$C = \sup_{\{X_i\}} \bar{h}(Y) - \bar{h}(Z) \qquad (15)$$

where the supremum is over all stationary Gaussian processes $\{X_i\}$ of the form $X_i = \sum_{k=1}^{\infty} b_k Z_{i-k}$ such that $\mathsf{E}(X_i^2) \leq P$.

We show that the capacity expression (13) can be rewritten in terms of the instability which can be tolerated under the power budget. Let the *sensitivity function* [30] be the transfer function from the output to the noise,

$$S(z) = \frac{Y(z)}{Z(z)}. \qquad (16)$$

By discrete version of Bode's integral [4] we know that if the closed loop is stable then

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \log(|S(e^{j\omega})|) d\omega = \sum_{j=1}^{m} \log(|\beta_j|) = I. \qquad (17)$$

Let $S_Y(e^{j\omega})$ and $S_Z(e^{j\omega})$ be the power spectral density of the output and the noise sequence, respectively. From (16) we have

$$|S(e^{j\omega})|^2 = \frac{S_Y(e^{j\omega})}{S_Z(e^{j\omega})} \qquad (18)$$

and plugging (18) into (17) we get

$$I = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log \left(\frac{S_Y(e^{j\omega})}{S_Z(e^{j\omega})} \right) d\omega. \qquad (19)$$

The entropy rate of a stationary Gaussian process $\{Y_i\}_{i=1}^{\infty}$ was shown by Kolmogorov [28] to be

$$\bar{h}(Y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{2} \log(2\pi e S_Y(e^{j\omega})) d\omega. \quad (20)$$

Plugging (20) into (15) and comparing with (19) we have

$$C = \sup \ I$$

where the supremum is over all causal $B(z)$ such that (14) is satisfied. On the other hand, $B(z) = \frac{X(z)}{Z(z)}$ can be written in terms of $F(z) = \frac{X(z)}{Y(z)}$ as follows,

$$B(z) = \frac{F(z)}{1 - F(z)}.$$

Note that $F(z)$ being causal is equivalent to $B(z)$ being causal and vice-versa. Therefore, considering constraint (14), the supremum can be taken over all causal filters $F(z)$ such that

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{F(e^{j\omega})}{1 - F(e^{j\omega})} \right|^2 S_Z(e^{j\omega}) d\omega \leq P.$$

∎

Theorem 2 states that the capacity of the stationary Gaussian channel represents the maximum instability of an LTI system which can be stabilized over the Gaussian channel by a linear controller with power at most P.

Next, we consider some special cases of the noise spectrum.

A. ARMA(1) Noise

Let the noise process be first-order autoregressive moving-average ARMA(1), i.e.,

$$S_Z(e^{j\omega}) = \frac{|1 + \alpha e^{j\omega}|}{|1 + \beta e^{j\omega}|} \quad (21)$$

where $\alpha \in [-1, 1]$ and $\beta \in (-1, 1)$.

Corollary 2: For the stationary Gaussian channel with noise process ARMA(1) given in (21), and power constraint P, we have

$$C_{MMSE} = C = \sup I \quad (22)$$

where the supremum is over all causal $B(z)$ such that the power constraint (14) is satisfied.

Proof: To show the first equality in (22), i.e., $C_{MMSE} = C$, note that the feedback capacity of the Gaussian channel with ARMA(1) noise is achieved [3, Theorem 5.3] by the Schalkwijk-Kailath scheme with probability of error going to zero doubly exponentially fast. Therefore, the sufficient condition (11) provided in Section III is satisfied and hence $C_{MMSE} = C$. The second equality in (22) follows from Theorem 2. ∎

Remark 4: Note that the AWGN channel, for which the noise spectrum is white, is a special case of ARMA(1) with $\alpha = \beta = 0$.

B. White Noise

The following corollary considers the special case of AWGN channels.

Corollary 3: For AWGN channels where the noise is i.i.d. and Gaussian, $Z_i \sim N(0, 1)$, the transfer function $F(z)$

$$F(z) = -(\beta^2 - 1) \frac{\beta^{-1} z^{-1}}{1 - \beta z^{-1}},$$

which has one pole at $\beta = \sqrt{1 + P} > 1$, achieves the MMSE capacity

$$C_{MMSE}(P) = C(P) = \frac{1}{2} \log(1 + P)$$

under power constraint P.

Proof: Note that for the AWGN channel, $S_Z(e^{j\omega}) = 1$, and by Corollary 2 we have

$$C_{MMSE}(P) = C(P) = \frac{1}{2} \log(1 + P).$$

Moreover, for $\beta = \sqrt{1 + P}$, the transfer function $F(z)$ has instability $I = \log(\beta)$, hence

$$C_{MMSE}(P) = C(P) = I.$$

It is left to show that $F(z)$ also satisfies the power constraint (12), i.e.,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} |B(e^{j\omega})|^2 S_Z(e^{j\omega}) d\omega \leq P. \quad (23)$$

where

$$B(z) = \frac{F(z)}{1 - F(z)} = -(\beta^2 - 1) \frac{\beta^{-1} z^{-1}}{1 - \beta^{-1} z^{-1}}.$$

Note that $B(z)$ is same as the optimal filter given in [3, Lemma 4.1] for the AWGN channel and similarly for $\beta > 1$ we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{1}{1 - \beta^{-1} e^{-j\omega}} \right|^2 d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \sum_{k=0}^{\infty} \beta^{-k} e^{-jk\omega} \right|^2 d\omega$$

$$= \sum_{k=0}^{\infty} \beta^{-k}$$

$$= \frac{1}{1 - \beta^{-2}}.$$

Hence,

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{F(e^{j\omega})}{1 - F(e^{j\omega})} \right|^2 S_Z(e^{j\omega}) d\omega$$

$$= (\beta^2 - 1)^2 \frac{\beta^{-2}}{1 - \beta^{-2}}$$

$$= (\beta^2 - 1)$$

$$= P \quad (24)$$

and $F(z)$ satisfy the power constraint. ∎

Finally, we show that the filter in Corollary 3 corresponds to the Schalkwijk and Kailath scheme [31], [32]. The transfer function $F(z) = \frac{X(z)}{Y(z)}$ in Corollary 3 can be represented in the time domain as follows,

$$X_i - \beta X_{i-1} = -\frac{\beta^2 - 1}{\beta} Y_{i-1}$$

or

$$X_i = \beta \left(X_{i-1} - \frac{\beta^2 - 1}{\beta^2} Y_{i-1} \right). \tag{25}$$

As it was shown earlier, for the transfer function $F(z)$ in Corollary 3 we have $\mathsf{E}(X_i^2) = P$. Hence, $\mathsf{E}(Y_i^2) = 1 + P$ and

$$\frac{\beta^2 - 1}{\beta^2} = \frac{P}{1 + P}$$
$$= \frac{\mathsf{E}(X_{i-1} Y_{i-1})}{\mathsf{E}(Y_{i-1}^2)}. \tag{26}$$

Therefore, (25) can be written as

$$X_i = \beta \left(X_{i-1} - \frac{\mathsf{E}(X_{i-1} Y_{i-1})}{\mathsf{E}(Y_{i-1}^2)} Y_{i-1} \right),$$

which is the recursive representation of the Schalkwijk and Kailath encoder [33].

V. GAUSSIAN MULTIPLE ACCESS CHANNEL

In this section, we extend the control representation for the communication over point-to-point channels to the feedback communication of N senders and one receiver over the AWGN-MAC depicted in Fig. 2a. Each sender $j \in \{1, \ldots, N\}$ wishes to reliably transmit a message $M_j \in \mathcal{M}_j := (0, 1)$ in the unit interval to the receiver. At each time i, the output of the channel is

$$Y_i = \sum_{j=1}^{N} X_{ji} + Z_i$$

where $X_{ji} \in \mathbb{R}$ is the transmitted symbol by sender j at time i, $Y_i \in \mathbb{R}$ is the output of the channel, and $\{Z_i\}$ is a discrete-time zero-mean white Gaussian noise process with unit average power, i.e., $\mathsf{E}(Z_i^2) = 1$, independent of $M_1, \ldots, M_N$. We assume that output symbols are causally fed back to the sender and the transmitted symbol X_{ji} for sender j at time i can depend on both the message M_j and the previous channel output sequence $Y^{i-1} := \{Y_1, Y_2, \ldots, Y_{i-1}\}$. We define a n-code for the AWGN-MAC as

1) N encoders: each encoder j, $j = 1, \ldots, N$, is a set of encoding maps $f_{ji} : (\mathcal{M}_j, \mathcal{Y}^{i-1}) \to \mathcal{X}_j$, $i = 1, \ldots, n$, also known to the receiver, such that

$$X_{ji} = f_{ji}(M_j, Y^{i-1}) \tag{27}$$

   and

2) a decoder: a decoding map $\phi : \mathcal{Y}^n \to \mathcal{M}_1 \times \ldots \times \mathcal{M}_N$ which determines the estimates of the messages $\hat{M}_1, \ldots, \hat{M}_N$ given Y^n

$$(\hat{M}_1, \ldots, \hat{M}_N) = \phi(Y^n). \tag{28}$$

We assume that the message vector $\mathbf{M} := (M_1, \ldots, M_N) \sim \mathrm{Unif}\left((0,1)^N\right)$ is uniformly distributed in a N dimensional unit box and the performance measure is the set of mean square errors,

$$D_j^{(n)} := \mathsf{E}\left((M_j - \hat{M}_j(Y^n))^2 \right) \quad j = 1, \ldots, N.$$

The set of *MSE exponents* $(E_1, \ldots, E_N)$ is called achievable if there exists a sequence of n-codes such that for $j = 1, \ldots, N$,

$$\liminf_{n \to \infty} -\frac{1}{2n} \log(D_j^{(n)}) \geq E_j.$$

We say that a sequence of n-codes has *asymptotic powers* $(\bar{P}_1, \ldots, \bar{P}_N)$ if

$$\lim_{n \to \infty} \mathsf{E}(X_{jn}^2) = \bar{P}_j, \quad j \in \{1, \ldots, N\}.$$

Similar to the point-to-point case, the information-theoretic achievable rates [28] are defined based on a sequence of massage sets whose size depend on the block length n. Consider the discrete message sets $\mathcal{M}_{jn} := \{1, 2, \ldots, 2^{nR_j}\}, j = 1, \ldots, N$. For a n-code with messages $M_{jn} \sim \mathrm{Unif}(\mathcal{M}_n), j = 1, \ldots, N$ uniformly distributed over the set $\mathcal{M}_{jn}$, let the probability of error be

$$P_e^{(n)} := \mathsf{P}\{(M_1, \ldots, M_N) \neq (\hat{M}_1, \ldots, \hat{M}_N)\}.$$

The set of rates $(R_1, \ldots, R_N)$ is called achievable under block power constraints $(P_1, \ldots, P_N)$ if there exists a sequence of n-codes with messages $M_j \sim \mathrm{Unif}(\mathcal{M}_{j,n})$, such that for $i \in \{1, \ldots, n\}$, and $j \in \{1, \ldots, N\}$,

$$\sum_{i=1}^{n} \mathsf{E}(X_{ji}^2) \leq nP_j \tag{29}$$

and $P_e^{(n)} \to 0$ as $n \to \infty$. We refer to $R = \sum_{j=1}^{N} R_j$ as the *sum rate* of a given code.

The following lemma presents the connection between the achievable MSE exponents and the achievable rates.

Lemma 4: Let MSE exponents $(E_1, \ldots, E_N)$ with asymptotic powers $(\bar{P}_1, \ldots, \bar{P}_N)$ be achievable and

$$R_j < E_j, \quad \bar{P}_j < P_j$$

for $j = 1, \ldots, N$. Then the rate-tuple $(R_1, \ldots, R_N)$ is also achievable and the block power constraints $(P_1, \ldots, P_N)$ are asymptotically satisfied.

   Proof: Let the MSE exponents $(E_1, \ldots, E_N)$ be achievable. By a similar argument as in Lemma 2 we can see that the rates $(R_1, \ldots, R_N)$ are also achievable if

$$R_j < E_j, \quad j = 1, \ldots, N.$$

Moreover, if $\bar{P}_j = \lim_{n \to \infty} \mathsf{E}(X_{jn}^2), j = 1, \ldots, N$ and $\bar{P}_j < P_j$, then using the Cesáro sum we can see that the constraint

$$\frac{1}{n} \sum_{i=1}^{n} \mathsf{E}(X_{ji}^2) \leq P_j, \quad j \in \{1, \ldots, N\}$$

is satisfied for sufficiently large n. ∎

A. LQG Approach for the AWGN-MAC

In the following we show that codes can be designed for the AWGN-MAC with feedback, based on the LTI systems controlled over a point-to-point AWGN channel depicted in Fig. 2b. We refer to these codes as LTI codes for the AWGN-MAC.
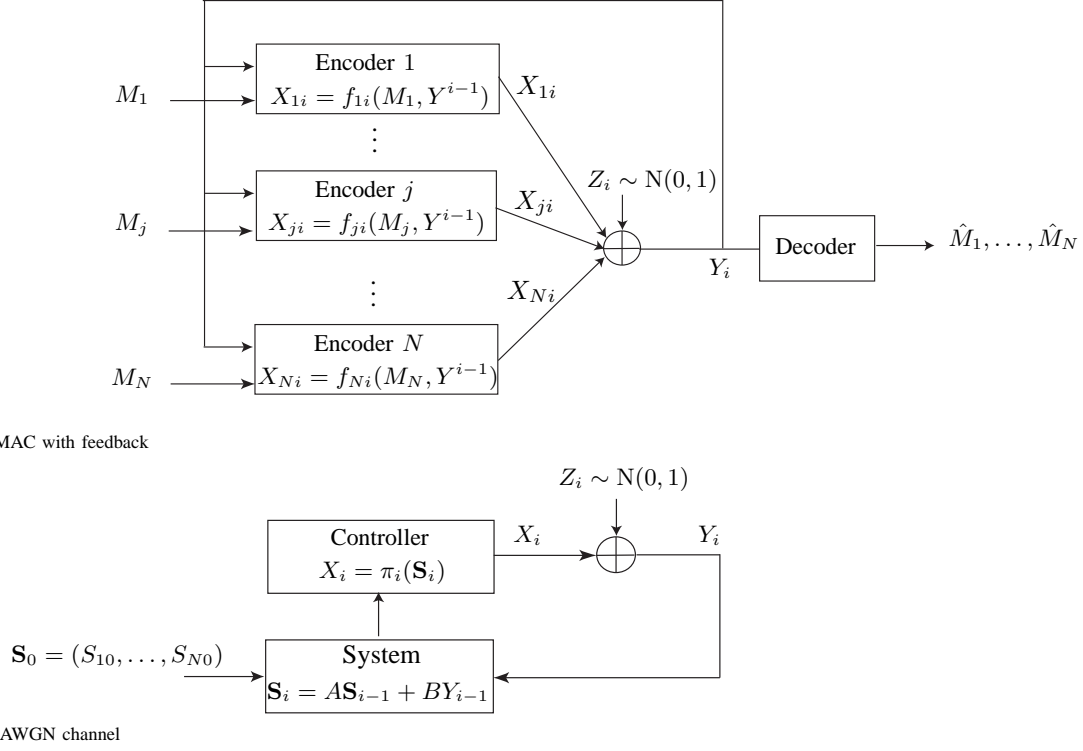
(a) The AWGN-MAC with feedback

(b) Control over AWGN channel

Fig. 2. The set of N encoders for feedback communication over the AWGN-MAC depicted in (a) can be represented as a dynamical system depicted in (b).

Let the state $\mathbf{S}_i \in \mathbb{C}^N$ be a complex vector of length N, and the system dynamics be

$$\mathbf{S}_i = A\mathbf{S}_{i-1} + BY_{i-1} \quad i = 1, \ldots, n \tag{30}$$

where

$$A = \begin{pmatrix} \beta_1\omega_1 & 0 & 0 & \ldots & 0 \\ 0 & \beta_2\omega_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & \beta_N\omega_N \end{pmatrix}, \ B = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{31}$$

such that $\beta_j > 1$ are real and $|\omega_j| = 1$ are distinct points on the unit circle of the complex plane. As it is clear from the system dynamics given in (30), the dependence of the state on the channel outputs is causal .

The controller is assumed to control each mode j separately and transmit a combined control signal (scalar) X_i as follows. It observes the current state $\mathbf{S}_i$ and picks a vector

$$\mathbf{X}_i := (X_{1i}, \ldots, X_{Ni}) \tag{32}$$

where

$$X_{ji} = \pi_{ji}(\mathbf{S}_i(j)), \quad j = 1, \ldots, N \tag{33}$$

can depend only on $\mathbf{S}_i(j)$, and then transmits the control signal

$$X_i = \sum_{j=1}^{N} X_{ji} = \sum_{j=1}^{N} \pi_{ji}(\mathbf{S}_i(j)). \tag{34}$$

For the purpose of analysis we assume complex transmissions over a complex channel,

$$Y_i = X_i + Z_i, \quad Z_i \sim \mathcal{CN}(0,1) \text{ i.i.d.} \tag{35}$$

where $X_i, Y_i \in \mathbb{C}$ are input and output of the channel, respectively, and $\{Z_i\}$ is a discrete-time zero-mean white complex Gaussian noise process with identity covariance. Note that one transmission over this complex channel can be viewed as two transmissions over the real channel. Hence, the achievable rates per each complex dimension are also achievable over the real channel.

We assume the complex messages $M_j \in \mathbb{C}$, $j = 1, \ldots, N$, in the AWGN-MAC are uniform over $(0,1) \times (0,1)$ and we set the initial state $\mathbf{S}_0 \in \mathbb{C}^N$ of the system as

$$\mathbf{S}_0 = \mathbf{M} = (M_1, \ldots, M_N).$$

Given the system (30) and the controller (34), we derive the LTI n-code for the AWGN-MAC as follows.

Definition 3: An LTI n-code for the AWGN-MAC based on LTI system (30) and (34) consists of

1) Encoding mappings: The encoder j recursively forms

$$\mathbf{S}_i(j) = \beta_1\omega_1\mathbf{S}_{i-1}(j) + Y_{i-1}, \ i = 1, \ldots, n \tag{36}$$

and at time i transmits $X_{ji} = \pi_{ji}(\mathbf{S}_i(j))$.

2) Decoding: At the end of the block, the decoder forms the estimate vector $\hat{\mathbf{M}}_n := -A^{-n}\hat{\mathbf{S}}_n$ where

$$\hat{\mathbf{S}}_i = A\hat{\mathbf{S}}_{i-1} + BY_{i-1}, \quad \hat{\mathbf{S}}_0 = 0, Y_0 = 0 \tag{37}$$

and picks $\hat{\mathbf{M}}_n(j)$ as the estimate of the message M_j.

Note that to design codes for the MAC it is necessary that the matrix A is diagonal as in (31) and that the control signal is of the form (34) since the encoders in the MAC are decentralized and do not have access to each other's messages. In the code described above, the dynamics of the j-th mode of the system $\mathbf{S}_i(j)$ represents the information based on which the encoder j picks the transmitted signal X_{ji} at time i. We say that the controller stabilizes the system if

$$\limsup_{n \to \infty} \mathsf{E}(|\mathbf{S}_n(j)|^2) < \infty, \quad j = 1, \ldots, N.$$

Lemma 5: If the controller of the form (34) stabilizes the linear system in (30), then the corresponding sequence of n-codes for the AWGN-MAC with feedback described in (36) and (37) achieves MSE exponents

$$E_j = \log(\beta_j), \quad j = 1, \ldots, N.$$

Proof: See Appendix D.

Remark 5: Note that the set $\{\beta_j\}$ depends only on the system[2] and not on the controller. This means that any stabilizing controller for the system (30) can be used to design a code for the AWGN-MAC, which achieves the same set of MSE exponents $(\log(\beta_1), \ldots, \log(\beta_N))$ or, according to Lemma 2, the same set of rates.

In the following, we assume that the system is fixed as in (30) and we find the stationary controller which minimizes the asymptotic power

$$\bar{P} := \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \mathsf{E}(|X_i|^2). \quad (38)$$

As we show below, the code for the AWGN-MAC which is based on this optimal controller is also optimal in terms of sum rate among the linear codes for the AWGN-MAC under equal power constraints.

For a controller of the form (34), let $K_{\mathbf{X}_i} := \mathrm{Cov}(\mathbf{X}_i)$ be the covariance of $\mathbf{X}_i$ given in (32). We say that $\bar{K}_{\mathbf{X}}$ is the *asymptotic covariance* if

$$\bar{K}_{\mathbf{X}} := \lim_{n \to \infty} K_{\mathbf{X}_n}.$$

Note that by (34), the asymptotic power of the controller can be written as $\bar{P} = \sum_{j,k=1}^{N} (\bar{K}_{\mathbf{X}})_{jk}$. Hence, $\bar{P}$ represents the asymptotic combined power of all the senders in the corresponding code for the AWGN-MAC, which also captures the correlation between the transmitted signals. Whereas, the asymptotic power of each sender in the AWGN-MAC is determined by $\bar{P}_j = (\bar{K}_{\mathbf{X}})_{jj}, j = 1, \ldots, N$.

Lemma 6: The optimal controller of the form (34) which minimizes the cost (38) is stationary and linear, i.e.,

$$X_i = -C\mathbf{S}_i,$$

where

$$C = (B'GB + 1)^{-1}B'GA \quad (39)$$

and G is the solution to the following discrete algebraic Riccati equation (DARE)

$$G = A'GA - A'GB(B'GB + 1)^{-1}B'GA. \quad (40)$$

[2]From now on, the word *system* refers to the the one given in (30).

Proof: Note that the optimal power $\bar{P}^*$ for the controllers of the form (34) can be lower bounded as

$$\bar{P}^* \geq \bar{P}_{\min}$$

where $\bar{P}_{\min}$ is the optimal cost considering a more general controller of the form

$$X_i = \pi_i(\mathbf{S}_i) \quad (41)$$

which picks the scalar action X_i based on the complete state $\mathbf{S}_i$ and is not necessarily separated for different modes as in (34). Considering the general control (41) we have a linear Gaussian quadratic control problem. From the theory of LQG [34] we know that the optimal control is linear, i.e.,

$$X_i = -C_i\mathbf{S}_i, \quad C_i = [c_{1i}c_{2i}\ldots c_{Ni}]. \quad (42)$$

and the stationary linear control $C = \lim_{i \to \infty} C_i$, which minimizes the asymptotic cost (38), is given by C in (39). Since the solution to this LQG problem given in (42) is of the form (34) we conclude that

$$\bar{P}^* = \bar{P}_{\min}$$

and the optimal controller of the form (34) is same as the optimal controller of the LQG problem. ∎

By Lemma 6, the optimal controller of the form (34) is linear. The following theorem provides rate and power analysis for the AWGN-MAC code based on a general stationary linear control of the form

$$X_i = -C\mathbf{S}_i, \quad C = [c_1c_2\ldots c_N]. \quad (43)$$

Theorem 3: Consider the stationary linear controller (43) for the system (30). If $A - BC$ is stable, i.e., the eigenvalues of $A - BC$ are inside the unit circle, then the corresponding code for AWGN-MAC designed based on this linear control achieves the rates

$$R_j < E_j = \log(\beta_j), \quad j = 1, \ldots, N$$

with asymptotic powers

$$\bar{P}_j = c_j^2 \bar{K}_{jj}, \quad j = 1, \ldots, N \quad (44)$$

where $\bar{K}$ is the unique solution of the following discrete algebraic Lyapunov equation (DALE)

$$\bar{K} = (A - BC)\bar{K}(A - BC)' + Q, \quad Q = BB'. \quad (45)$$

Proof: First note that by Lemma 5, MSE exponents $E_j = \log(\beta_j), j = 1, \ldots, N$, are achievable and hence, by Lemma 4, any rate $R_j < E_j, j = 1, \ldots, N$ is achievable in each complex dimension.

To find the asymptotic powers, note that given a stationary linear controller of the form (43), the closed loop system given (30) can be written as

$$\mathbf{S}_i = (A - BC)\mathbf{S}_{i-1} + BZ_{i-1}. \quad (46)$$

Consider the code corresponding to the linear control (43), where the encoder j transmits

$$X_{ji} = c_j\mathbf{S}_i(j)$$

at time i and let $K_i := \text{Cov}(\mathbf{S}_i)$. Then, the asymptotic power $\bar{P}_j$ is

$$\bar{P}_j = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mathsf{E}(X_{ji}^2) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n |c_j|^2 (K_i)_{jj}. \quad (47)$$

From the closed loop dynamic given in (46) we have the following Lyapunov recursion,

$$K_n = (A - BC)K_{n-1}(A - BC)' + Q, \quad Q = BB'.$$

If $A - BC$ is stable, we know [34] that

$$\lim_{n\to\infty} K_n = \bar{K} \succ 0, \quad (48)$$

where $\bar{K}$ is the unique positive-definite solution of the corresponding discrete algebraic Lyapunov equation (DALE)

$$\bar{K} = (A - BC)\bar{K}(A - BC)' + Q.$$

Therefore, by (48) and the Cesáro mean theorem, the asymptotic power in (47) becomes $\bar{P}_j = |c_j|^2 \bar{K}_{jj}$. ∎

B. Kramer Code vs. Optimal LTI Code

Now consider the symmetric system

$$\omega_j = e^{\frac{2\pi\sqrt{-1}}{N}(j-1)}$$
$$\beta_j = \beta \quad j = 1,\ldots,N. \quad (49)$$

and let the *optimal LTI code* for the AWGN-MAC be the code based on the optimal control given in Lemma 6. The following lemma characterizes the matrix G for the symmetric choice of matrix A given in (49).

Lemma 7: Let $\beta_j = \beta$ and $\omega_j = e^{2\pi\sqrt{-1}\frac{(j-1)}{N}}$. The unique positive-definite solution $G \succ 0$ to the DARE (40) is circulant with real eigenvalues satisfying $\lambda_i = \frac{1}{\beta^2}\lambda_{i-1}$ for $i = 2,\ldots,N$. The largest eigenvalue λ_1 satisfies

$$1 + N\lambda_1 = \beta^{2N} \quad (50)$$

$$\left(1 + \lambda_1\left(N - \frac{\lambda_1}{G_{11}}\right)\right) = \beta^{2(N-1)}. \quad (51)$$

Proof: See Appendix E.

The following theorem shows that under equal power constraints for all the senders, the optimal LTI code for the system (49) corresponds to the linear code proposed by Kramer [7], which is known to be sum rate optimal within the class of generalized feedback codes [8].

Theorem 4: Let $R(P)$ be the sum rate achieved by Kramer's code with equal power P. There exists a $\beta > 1$ such that the optimal LTI code based on the system (49) achieves the sum rate $R(P)$ and satisfies the equal power constraint P asymptotically.

Proof: By Theorem 3, the optimal LTI code based on the symmetric system (49) and the optimal controller in (39) achieves the symmetric MSE exponent E, i.e, $E_j = E = \log(\beta)$, $j = 1,\ldots,N$ with asymptotic powers $\bar{P}_j = c_j^2 \bar{K}_{jj}$, $j = 1,\ldots,N$ where $\bar{K}$ is given in (45). Moreover, by Lemma 4, rates $R_j < \log(\beta_j)$, $j = 1,\ldots,N$ are achievable and therefore the sum rate

$$R < N\log(\beta) \quad (52)$$

is achievable.

On the other hand, let $R(P)$ denote the sum rate achievable by Kramer's code under symmetric per-symbol power P. From [7] we have

$$R(P) < \frac{1}{2}\log(1 + NP\phi(P)) \quad (53)$$

where $\phi(P) \in \mathbb{R}$ is the unique solution in the interval $[1, N]$ to

$$(1 + NP\phi)^{N-1} = (1 + P\phi(N - \phi))^N. \quad (54)$$

Therefore, by picking

$$R(P) < N\log(\beta) < \log(1 + NP\phi(P)) \quad (55)$$

the LTI code can achieve sum rate equal to $R(P)$.

It remains to show that the asymptotic powers in the optimal LTI code satisfy

$$\bar{P}_j < P, \quad j = 1,\ldots,N. \quad (56)$$

Towards this end, we first show that $\bar{P}_j = G_{jj}$, where G is the solution to (40). Note that the DARE given in (40) can be equivalently written as the following discrete algebraic Lyapunov equation (DALE)

$$G = (A - BC)'G(A - BC) + C'C \quad (57)$$

and compared with (45), the diagonal elements of G and K can be related as follows

$$G_{jj} = c_j^2 \bar{K}_{jj}. \quad (58)$$

Besides, from (44) we know $\bar{P}_j = c_j^2 \bar{K}_{jj}$. Hence, we conclude that the diagonal elements of G represents the asymptotic powers in the optimal LTI code,

$$\bar{P}_j = G_{jj}. \quad (59)$$

Now we will show that if β satisfies (55), then $G_{jj} < P$ for the corresponding G. First, note that from (50) and (51) we have

$$\left(1 + N\lambda_1\right)^{N-1} = \left(1 + \lambda_1(N - \frac{\lambda_1}{G_{11}})\right)^N. \quad (60)$$

Comparing with (54) and noting that G is circulant we get

$$\lambda_1 = G_{jj}\phi(G_{jj}), \quad j = 1,\ldots,N. \quad (61)$$

Now if (55) holds, from (50) we know

$$\frac{1}{2}\log(1 + N\lambda_1) = N\log(\beta) < \frac{1}{2}\log(1 + NP\phi(P))$$

and hence

$$\lambda_1 < P\phi(P). \quad (62)$$

From (61) and (62) we have $G_{jj}\phi(G_{jj}) < P\phi(P)$, and $G_{jj} < P$ follows from monotonicity of $\phi(P)$ in P. Combined with (59), then (56) follows. ∎

VI. CONCLUSION

Communication and control problems have different goals, but they both deal with information dynamics and in many cases they share similar formulations which can be tackled by tools and techniques developed in both fields. Understanding the interface between these two theories has become even more important in the last decade, as we have witnessed technological advancements leading to the convergence of computing, communication, and control over networked platforms of embedded systems. This paper attempted, *ad rem*, one step in this direction.

APPENDIX A
PROOF OF LEMMA 1

Note that without loss of generality we can decompose the encoding functions f_i in (1) into two steps as follows. First, based on the past output Y^{i-1}, the encoder picks a function

$$\tilde{f}_i : \mathcal{M} \to X. \tag{63}$$

Then it transmits

$$X_i = \tilde{f}_i(M).$$

With this separation, the encoder (controller) can be defined by the set of mappings

$$\tilde{\pi}_i : \mathcal{Y}^{i-1} \to \{\tilde{f}\}$$

where $\{\tilde{f}\}$ is the set of all functions given in (63). Note that this new encoder $\{\tilde{\pi}_i\}_{i=1}^n$ only observes the past output Y^{i-1} and not the message M. Therefore, one can view the feedback communication as a control problem where the state is M and the controller has partial observations Y^{i-1}. Based on standard results in stochastic control [35, Chapter 6], there is no loss of optimality if the mapping $\tilde{\pi}_i$ is picked only according to the conditional distribution $F_{M|Y^{i-1}}(\cdot|Y^{i-1})$. Therefore, to find X_i it is sufficient to have the message M and its posterior distribution $F_{M|Y^{i-1}}(\cdot|Y^{i-1})$.

APPENDIX B
PROOF OF LEMMA 2

Consider a sequence of n-codes for $M \in (0,1)$ such that

$$E \leq \liminf_{n \to \infty} -\frac{1}{2n} \log(D^{(n)}). \tag{64}$$

We show that a rate $R < E$ is achievable. Consider

$$\mathsf{P}(|M - \hat{M}| > \frac{1}{2} \cdot 2^{-nR}) \leq 4 \cdot 2^{2nR} \cdot D^{(n)} \tag{65}$$

$$\leq 4 \cdot 2^{2nR} \cdot 2^{-2n(E-\epsilon_n)} \tag{66}$$

$$= 4 \cdot 2^{-2n(E-R-\epsilon_n)} \tag{67}$$

for some ϵ_n where $\epsilon_n \to 0$ as $n \to \infty$, where (65) follows from Chebyshev inequality and (66) follows from (64). If $R < E$, from (67) we have

$$\mathsf{P}(|M - \hat{M}| > \frac{1}{2} \cdot 2^{-nR}) \to 0$$

as $n \to \infty$ and the decoder can pick intervals Δ_n such that

$$|\Delta_n| = 2^{-nR}$$

and

$$\mathsf{P}(M \notin \Delta_n) \to 0 \text{ as } n \to \infty. \tag{68}$$

By Lemma II.3 in [27], if condition (68) is satisfied then for all n, we can map the message set $\mathcal{M}_n = \{1, \ldots, 2^{nR}\}$ into a set of message points $\tilde{\mathcal{M}}_n \in (0,1)$ in the unit interval such that the minimum distance between the two message points is 2^{-nR}. Therefore, using the interval decoder described above which satisfy (68), $P_e^{(n)} \to 0$ as $n \to \infty$ and rate R is achievable.

To complete the proof, let rate R be achievable by a given sequence of n-codes such that $\lim_{n\to\infty} \frac{P_e^{(n)}}{2^{-2nR}} \to 0$. For each n, we divide the unit interval $(0,1)$ into 2^{-nR} equal sub-intervals and map the continuous message $M \in (0,1)$ to the discrete message $M_n \in \mathcal{M}_n = \{1, \ldots, 2^{nR}\}$ according to the sub-interval M lies in. To communicate M, we send the corresponding M_n using the given n-code, and we pick the middle point of the interval corresponding to the decoded message $\hat{M}_n$ as the estimate of $M \in (0,1)$. The MSE for M can be (loosely) upper bounded by 2^{-2nR} if the message M_n is decoded correctly, and by 1 in case of an error. Hence,

$$D^{(n)} \leq P_e^{(n)} + (1 - P_e^{(n)})2^{-2nR} \tag{69}$$

From (69) and the assumption $\lim_{n\to\infty} \frac{P_e^{(n)}}{2^{-2nR}} \to 0$ we have

$$\liminf_{n \to \infty} -\frac{1}{2n} D^{(n)} \geq R$$

and the MSE exponent $E = R$ is achievable.

APPENDIX C
PROOF OF LEMMA 3

We show that

$$h(Y^n) \geq h(Z^n) + I(S_0; Y^n)$$

with equality if and only if X_i is a function of (S_0, Y^{i-1}) for $i = 1, \ldots, n$.

Consider

$$h(Y^n) = h(Y^n|S_0) + I(S_0; Y^n)$$

$$= \sum_{i=1}^n h(Y_i|Y^{i-1}, S_0) + I(S_0; Y^n)$$

$$\geq \sum_{i=1}^n h(Y_i|Y^{i-1}, X^i, S_0) + I(S_0; Y^n) \tag{70}$$

$$= \sum_{i=1}^n h(Y_i, Z_i|Y^{i-1}, X^i, S_0) + I(S_0; Y^n) \tag{71}$$

$$= \sum_{i=1}^n h(Z_i|Y^{i-1}, X^i, S_0) + I(S_0; Y^n)$$

$$= \sum_{i=1}^n h(Z_i|Y^{i-1}, X^i, S_0, Z^{i-1}) + I(S_0; Y^n) \tag{72}$$

$$= \sum_{i=1}^n h(Z_i|Z^{i-1}) + I(S_0; Y^n) \tag{73}$$

$$= h(Z^n) + I(S_0; Y^n) \tag{74}$$

where

- the inequality (70) comes from the fact that conditioning reduces entropy, and equality holds iff $Y_i \rightarrow (S_0, Y^{i-1}) \rightarrow X_i$ form a Markov chain for $i = 1, \ldots, n$, and since $Y_i = X_i + Z_i$ this Markov chain holds iff X_i is a deterministic function of (S_0, Y^{i-1}),
- the equality (71) and (72) follows from the fact that the channel is invertible, and
- the equality (73) follows since $Z_i \rightarrow Z^{i-1} \rightarrow (Y^{i-1}, X^i, S_0)$ form a Markov chain.

APPENDIX D
PROOF OF LEMMA 5

The system dynamics given in (30) can be rewritten as

$$\mathbf{S}_i = A\mathbf{S}_{i-1} + BY_{i-1}$$
$$= A^i\mathbf{S}_0 + \hat{\mathbf{S}}_i \tag{75}$$

where $\hat{\mathbf{S}}_i$ is given in (37). Plugging $i = n$ and $\mathbf{S}_0 = \mathbf{M}$ into (75) and multiplying both sides by A^{-n} we have

$$A^{-n}\mathbf{S}_n = \mathbf{M} + A^{-n}\hat{\mathbf{S}}_n$$
$$= \mathbf{M} - \hat{\mathbf{M}}_n$$

where $\hat{\mathbf{M}} = -A^{-n}\hat{\mathbf{S}}_n$ is the estimate of the message (see Definition 3). The covariance matrix of the error vector $\mathbf{e}_n := \mathbf{M} - \hat{\mathbf{M}}_n = A^{-n}\mathbf{S}_n$ can be written as

$$\text{Cov}(\mathbf{e}_n) = A^{-n}K_n A'^{-n}$$

where $K_n := \text{Cov}(\mathbf{S}_n)$ is the covariance matrix of $\mathbf{S}_n$. Therefore, the MSE for the sender j is

$$D_j^{(n)} = \mathsf{E}\left(|\mathbf{e}_n(j)|^2\right) = \beta_j^{-2n}(K_n)_{jj}. \tag{76}$$

By the assumption of stability $\limsup_{n\to\infty}(K_n)_{jj} < \infty$ and from (76) we have

$$\liminf_{n\to\infty} -\frac{1}{2n}\log(D_j^{(n)}) = \log(\beta_j).$$

Hence, the MSE exponents $E_j = \log(\beta_j)$ for $j = 1, \ldots, N$ are achievable.

APPENDIX E
PROOF OF LEMMA 7

Let A and B be of the form (31) with symmetric parameters $\beta_j = \beta$ and $\omega_j = e^{2\pi\sqrt{-1}\frac{(j-1)}{N}}$. Then we show that the unique positive-definite solution $G \succ 0$ the following discrete algebraic Riccati equation (DARE)

$$G = A'GA - A'GB(B'GB + 1)^{-1}B'GA \tag{77}$$

is circulant with real eigenvalues satisfying $\lambda_i = \frac{1}{\beta^2}\lambda_{i-1}$ for $i = 2, \ldots, N$, and the largest eigenvalue λ_1 satisfies

$$1 + N\lambda_1 = \beta^{2N}$$
$$\left(1 + \lambda_1\left(N - \frac{\lambda_1}{G_{11}}\right)\right) = \beta^{2(N-1)}.$$

Note that any circulant matrix can be written as $Q\Lambda Q'$, where Q is the N point DFT matrix with

$$Q_{jk} = \frac{1}{\sqrt{N}}e^{-2\pi\sqrt{-1}(j-1)(k-1)/N} \tag{78}$$

and $\Lambda = \text{diag}([\lambda_1, \ldots, \lambda_N])$ is the matrix with eigenvalues on its diagonal. We show that the circulant matrix $\tilde{G} = Q\Lambda Q'$ with positive $\lambda_j > 0$, such that $\lambda_j = \lambda_{j-1}/\beta^2$ for $j \geq 2$, satisfies the Riccati equation (77). Plugging $Q\Lambda Q'$ into (77) and rearranging we get

$$\Lambda = (Q'AQ)\Lambda(Q'AQ)' - ((Q'AQ) \Lambda (Q'B))$$
$$(1 + B'Q\Lambda Q'B)^{-1}((Q'AQ) \Lambda (Q'B))'.$$

For this symmetric choice of A we have

$$Q'AQ = \beta \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \end{pmatrix}, Q'B = \begin{pmatrix} \sqrt{N} \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Hence,

$$(Q'AQ)\Lambda(Q'AQ)' = \beta^2 \begin{pmatrix} \lambda_2 & 0 & \ldots & 0 \\ 0 & \lambda_3 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_1 \end{pmatrix}$$

$$(Q'AQ)\Lambda(Q'B) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \beta\lambda_1\sqrt{N} \end{pmatrix}$$

and the Riccati equation is transformed into N diagonal equations. The first $N-1$ equations are

$$\lambda_j = \beta^2\lambda_{j+1}, \quad j = 1, \ldots, N-1 \tag{79}$$

and the N-th equation is

$$\lambda_N = \beta^2\lambda_1 - \frac{\beta^2\lambda_1^2 N}{1 + N\lambda_1}. \tag{80}$$

From (79) we see that λ_1 is the largest eigenvalue and $\lambda_N = \beta^{-2(N-1)}\lambda_1$. Combining with (80) we get

$$(1 + N\lambda_1) = \beta^{2N}. \tag{81}$$

Hence, λ_1 is real and so are $\lambda_j, j = 2, \ldots, N$.

On the other hand, we consider the diagonal equations of the original DARE in (77). First, note that from the form of Q in (78), $\lambda_1 = \sigma_1$ where

$$\sigma_j := \sum_{k=1}^{N} \tilde{G}_{jk}.$$

Moreover, since $\tilde{G}$ is circulant we know that $\sigma_j = \sigma_1$ and $G_{jj} = G_{11}$ for all $j = 1, \ldots, N$, and $(1+B'\tilde{G}B) = 1+N\lambda_1$. Hence, the diagonal equations of (77), i.e.,

$$\tilde{G}_{jj} = \beta^2\tilde{G}_{jj} - \beta^2\frac{\sigma_j^2}{(1 + B'\tilde{G}B)}, \quad j = 1, \ldots, N \tag{82}$$

are equivalent to

$$\beta^2 = \frac{1 + N\lambda_1}{1 + \lambda_1\left(N - \frac{\lambda_1}{\tilde{G}_{11}}\right)} \tag{83}$$

and by (81) we have

$$\left(1 + \lambda_1\left(N - \frac{\lambda_1}{\tilde{G}_{jj}}\right)\right) = \beta^{2(N-1)}. \tag{84}$$

Combining (79), (81), and (84) completes the proof.

ACKNOWLEDGMENTS

REFERENCES

[1] H. S. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proceedings of the IEEE*, vol. 59, pp. 1557–1566, 1971.

[2] N. Elia, "When Bode meets Shannon: Control oriented feedback communication schemes," *IEEE Trans. Automatic Control*, vol. 47, pp. 1477–1488, 2004.

[3] Y.-H. Kim, "Feedback capacity of stationary Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 56, pp. 57–85, 2010.

[4] H. K. Sung and S. Hara, "Properties of sensitivity and complementary sensitivity functions in single input and single output digital control systems," *Int. J. Control*, vol. 48, pp. 2429–2439, 1988.

[5] H. W. Bode, *Network Analysis and Feedback Amplifier Design*. Princeton, NJ: Van Nostrad, 1945.

[6] J. S. Freudenberg and D. P. Looze, *Frequency Domain Properties of Scalar and Multivariable Systems*. New York: Springer, 1988.

[7] G. Kramer, "Feedback strategies for white Gaussian interference networks," *IEEE Trans. Inf. Theory*, vol. 48, pp. 1423–1438, 2002.

[8] E. Ardestanizadeh, M. A. Wigger, T. Javidi, and Y. H. Kim, "Linear sum capacity for Gaussian multiple access channel with feddback," 2010, preprint, http://arxiv.org/abs/1002.1781.

[9] L. H. Ozarow, "The capacity of the white Gaussian multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, pp. 623–629, 1984.

[10] T. E. Duncan, "On the calculation of mutual information," *SIAM J. Appl. Math.*, vol. 19, p. 215220, 1970.

[11] G. D. Forney, "On the role of MMSE estimation in approaching the information-theoretic limits of linear Gaussian channels: Shannon meets Wiener," in *Proceedings of the 41th Annual Allerton Conference on Communication, Control, and Computation, Monticello, Illinois*, 2003.

[12] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1261–1282, 2005.

[13] M. Zakai, "On mutual information, likelihood ratios, and estimation error for the additive Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 51, pp. 3017–3024, 2005.

[14] J. Liu and N. Elia, "Convergence of fundamental limitations in feedback communication, estimation, and feedback control over Gaussian channels," 2009, preprint, http://arxiv.org/abs/0910.0320.

[15] S. Mitter and N. Newton, "A variational approach to nonlinear estimation," *SIAM J. Control Optim.*, vol. 42, pp. 1813–1833, 2003.

[16] ——, "Information and entropy flow in the Kalman–Bucy filter," *J. Stat. Phys.*, vol. 118, pp. 145–176, 2005.

[17] P. A. Iglesias, "An analogue of Bode's integral for stable non-linear systems: Relations to entropy," in *IEEE Conference on Decision and Control (CDC)*, 2001, pp. 3419–3420.

[18] G. Zang and P. A. Iglesias, "Nonlinear extension of Bode's integral based on an information theoretic interpretation," *Systems and Control Letters*, pp. 11–29, 2003.

[19] N. C. Martins, M. A. Dahleh, and J. C. Doyle, "Fundamental limitations of disturbance attenuation in the presence of side information," *IEEE Trans. Automatic Control*, vol. 52, pp. 56–66, 2007.

[20] N. C. Martins and M. A. Dahleh, "Feedback control in the presence of noisy channels: Bode-like fundamental limitations of performance," *IEEE Trans. Automatic Control*, vol. 53, pp. 1604–1615, 2008.

[21] S. Yu and P. G. Mehta, "Bode-like fundamental performance limitations in control of nonlinear systems," 2008, preprint, http://mechse.illinois.edu/sites/index.php?id=442%7CPublications.

[22] W. Wu, S. Vishwanath, and A. Arapostathis, "Gaussian interference networks with feedback: Duality, sum capacity and dynamic team problem," in *Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computation, Monticello, Illinois*, 2005.

[23] S. Tatikonda and S. Mitter, "The capacity of channels with feedback," *IEEE Trans. Inf. Theory*, vol. 55, pp. 323–349, 2009.

[24] T. P. Coleman, "A stochastic control viewpoint on 'posterior matching-style' communication schemes," in *IEEE International Symposium on Information Theory (ISIT)*, 2009.

[25] S. K. Mitter, "Control with limited information," *European Journal of Control*, vol. 7, pp. 122–131, 2001.

[26] H. S. Witsenhausen, "A counterexample in stochastic optimum control," *Siam Journal on Control and Optimization*, vol. 6, pp. 131–147, 1968.

[27] O. Shayevitz and M. Feder, "Optimal feedback communication via posterior matching," preprint, http://arxiv.org/abs/0909.4828.

[28] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. 2nd ed. New York: Wiley, 2006.

[29] T. M. Cover and S. Pombra, "Gaussian feedback capacity," *IEEE Trans. Inf. Theory*, vol. 35, pp. 37–43, 1989.

[30] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum, *Feedback Control Theory*. New York: Macmillan, 1992.

[31] J. P. M. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandlimited constraint," *IEEE Trans. Inf. Theory*, vol. 12, pp. 172–182, 1966.

[32] J. P. M. Schalkwijk, "A coding scheme for additive noise channels with feedback–II: Band-limited signals," *IEEE Trans. Inf. Theory*, vol. 12, pp. 183–189, 1966.

[33] A. El Gamal and Y.-H. Kim, *Lecture notes on network information theory*. Stanford University and University of California San Diego, 2009.

[34] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*. 2nd ed. New York: Wiley, 2006.

[35] P. R. Kumar and P. Varaiya, *Stochastic Control: Estimation, Identification, and Adaptive Control*. Prentice Hall, 1986.